

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Choi et. al.

(ii) TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines

(iii) NUMBER OF SEQUENCES: 452

(iv) CORRESPONDENCE ADDRESS:

- (A) ADDRESSEE: Human Genome Sciences, Inc.
- (B) STREET: 9410 Key West Avenue
- (C) CITY: Rockville
- (D) STATE: Maryland
- (E) COUNTRY: USA
- (F) ZIP: 20850

(v) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
- (B) COMPUTER: HP Vectra 486/33
- (C) OPERATING SYSTEM: MSDOS version 6.2
- (D) SOFTWARE: ASCII Text

(vi) CURRENT APPLICATION DATA:

- (A) APPLICATION NUMBER: Unassigned
- (B) FILING DATE: Herewith
- (C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: 08/961,083
- (B) FILING DATE: OCT-30-1997

(viii) ATTORNEY/AGENT INFORMATION:

- (A) NAME: Michelle S. Marks
- (B) REGISTRATION NUMBER: 41,971
- (C) REFERENCE/DOCKET NUMBER: PB340P2C1

(vi) TELECOMMUNICATION INFORMATION:

- (A) TELEPHONE: (301) 309-8504
- (B) TELEFAX: (301) 309-8512

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1999 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double

09765972 0422001

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

TAAATCTAC GACAATAAAA ATCAACTCAT TGCTGACTTG GGTCTGAAC GCCGCGTCAA	60
TGCCCCAAGCT AATGATATTC CCACAGATTT GGTAAAGGCA ATCGTTTCTA TCGAAGACCA	120
TCGCTTCTTC GACCACAGGG GGATTGATAC CATCCGTATC CTGGGAGCTT TCTTGCGCAA	180
TCTGCAAAGC AATTCCCTCC AAGGTGGATC AACTCTCACC CAACAGTTGA TTAAGTTGAC	240
TTACTTTTCA ACTTCGACTT CCGACCAGAC TATTTCTCGT AAGGCTCAGG AAGCTTGGTT	300
AGCGATTGAG TTAGAACAAA AAGCAACCAA GCAAGAAATC TTGACCTACT ATATAAATAA	360
GGTCTACATG TCTAATGGGA ACTATGGAAT GCAGACAGCA GCTCAAACT ACTATGGTAA	420
AGACCTCAAT AATTTAAGTT TACCTCAGTT AGCCTTGCTG GCTGGAATGC CTCAGGCACC	480
AAACCAATAT GACCCCTATT CACATCCAGA AGCAGCCCAA GACCGCCGAA ACTTGGTCTT	540
ATCTGAAATG AAAAATCAAG GCTACATCTC TGCTGAACAG TATGAGAAAG CAGTCAATAC	600
ACCAATTACT GATGGACTAC AAAGTCTCAA ATCAGCAAGT AATTACCCTG CTTACATGGA	660
TAATTACCTC AAGGAAGTCA TCAATCAAGT TGAAGAAGAA ACAGGCTATA ACCTACTCAC	720
AACTGGGATG GATGTCTACA CAAATGTAGA CCAAGAAGCT CAAAAACATC TGTGGGATAT	780
TTACAATACA GACGAATACG TTGCCTATCC AGACGATGAA TTGCAAGTCG CTTCTACCAT	840
TGTTGATGTT TCTAACGGTA AAGTCATTGC CCAGCTAGGA GCACGCCATC AGTCAAGTAA	900
TGTTTCCTTC GGAATTAACC AAGCAGTAGA AACAAACCGC GACTGGGGAT CAACTATGAA	960
ACCGATCACA GACTATGCTC CTGCCCTGGA GTACGGTGTC TACGATTCAA CTGCTACTAT	1020
CGTTCACGAT GAGCCCTATA ACTACCCTGG GACAAATACT CCTGTTTATA ACTGGGATAG	1080
GGGCTACTTT GGCAACATCA CCTTGCAATA CGCCCTGCAA CAATCGCGAA ACGTCCCAGC	1140
CGTGGAAGCT CTAAACAAGG TCGGACTCAA CCGCGCCAAG ACTTTCCTAA ATGGTCTAGG	1200
AATCGACTAC CCAAGTATTC ACTACTCAA TGCCATTTCA AGTAACACAA CCGAATCAGA	1260
CAAAAAATAT GGAGCAAGTA GTGAAAAGAT GGCTGCTGCT TACGCTGCCT TTGCAAATGG	1320
TGGAACTTAC TATAAACCAA TGTATATCCA TAAAGTCGTC TTTAGTGATG GGAGTGAAAA	1380
AGAGTTCTCT AATGTCGGAA CTCGTGCCAT GAAGGAAACG ACAGCCTATA TGATGACCGA	1440
CATGATGAAA ACAGTCTTGA CTTATGGAAC TGGACGAAAT GCCTATCTTG CTTGGCTCCC	1500
TCAGGCTGGT AAAACAGGAA CCTCTAACTA TACAGACGAG GAAATTGAAA ACCACATCAA	1560
GACCTCTCAA TTTGTAGCAC CTGATGAACT ATTTGCTGGC TATACGCGTA AATATTCAAT	1620
GGCTGTATGG ACAGGCTATT CTAACCGTCT GACACCACTT GTAGGCAATG GCCTTACGGT	1680
CGCTGCCAAA GTTTACCGCT CTATGATGAC CTACCTGTCT GAAGGAAGCA ATCCAGAAGA	1740

0976527-042204

TTGGAATATA CCAGAGGGGC TCTACAGAAA TGGAGAATTC GTATTTAAAA ATGGTGCTCG 1800
 TTCTACGTGG AACTCACCTG CTCCACAACA ACCCCCATCA ACTGAAAGTT CAAGCTCATC 1860
 ATCAGATAGT TCAACTTCAC AGTCTAGCTC AACCACTCCA AGCACAAATA ATAGTACGAC 1920
 TACCAATCCT AACAATAATA CGCAACAATC AAATACAACC CCTGATCAAC AAAATCAGAA 1980
 TCCTCAACCA GCACAACCA 1999

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 666 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Lys Ile Tyr Asp Asn Lys Asn Gln Leu Ile Ala Asp Leu Gly Ser Glu
 1 5 10 15
 Arg Arg Val Asn Ala Gln Ala Asn Asp Ile Pro Thr Asp Leu Val Lys
 20 25 30
 Ala Ile Val Ser Ile Glu Asp His Arg Phe Phe Asp His Arg Gly Ile
 35 40 45
 Asp Thr Ile Arg Ile Leu Gly Ala Phe Leu Arg Asn Leu Gln Ser Asn
 50 55 60
 Ser Leu Gln Gly Gly Ser Thr Leu Thr Gln Gln Leu Ile Lys Leu Thr
 65 70 75 80
 Tyr Phe Ser Thr Ser Thr Ser Asp Gln Thr Ile Ser Arg Lys Ala Gln
 85 90 95
 Glu Ala Trp Leu Ala Ile Gln Leu Glu Gln Lys Ala Thr Lys Gln Glu
 100 105 110
 Ile Leu Thr Tyr Tyr Ile Asn Lys Val Tyr Met Ser Asn Gly Asn Tyr
 115 120 125
 Gly Met Gln Thr Ala Ala Gln Asn Tyr Tyr Gly Lys Asp Leu Asn Asn
 130 135 140
 Leu Ser Leu Pro Gln Leu Ala Leu Leu Ala Gly Met Pro Gln Ala Pro
 145 150 155 160
 Asn Gln Tyr Asp Pro Tyr Ser His Pro Glu Ala Ala Gln Asp Arg Arg
 165 170 175
 Asn Leu Val Leu Ser Glu Met Lys Asn Gln Gly Tyr Ile Ser Ala Glu
 180 185 190
 Gln Tyr Glu Lys Ala Val Asn Thr Pro Ile Thr Asp Gly Leu Gln Ser
 195 200 205
 Leu Lys Ser Ala Ser Asn Tyr Pro Ala Tyr Met Asp Asn Tyr Leu Lys

09763272-012204

210	215	220
Glu Val Ile Asn Gln Val Glu Glu Glu Thr Gly Tyr Asn Leu Leu Thr 225 230 235 240		
Thr Gly Met Asp Val Tyr Thr Asn Val Asp Gln Glu Ala Gln Lys His 245 250 255		
Leu Trp Asp Ile Tyr Asn Thr Asp Glu Tyr Val Ala Tyr Pro Asp Asp 260 265 270		
Glu Leu Gln Val Ala Ser Thr Ile Val Asp Val Ser Asn Gly Lys Val 275 280 285		
Ile Ala Gln Leu Gly Ala Arg His Gln Ser Ser Asn Val Ser Phe Gly 290 295 300		
Ile Asn Gln Ala Val Glu Thr Asn Arg Asp Trp Gly Ser Thr Met Lys 305 310 315 320		
Pro Ile Thr Asp Tyr Ala Pro Ala Leu Glu Tyr Gly Val Tyr Asp Ser 325 330 335		
Thr Ala Thr Ile Val His Asp Glu Pro Tyr Asn Tyr Pro Gly Thr Asn 340 345 350		
Thr Pro Val Tyr Asn Trp Asp Arg Gly Tyr Phe Gly Asn Ile Thr Leu 355 360 365		
Gln Tyr Ala Leu Gln Gln Ser Arg Asn Val Pro Ala Val Glu Thr Leu 370 375 380		
Asn Lys Val Gly Leu Asn Arg Ala Lys Thr Phe Leu Asn Gly Leu Gly 385 390 395 400		
Ile Asp Tyr Pro Ser Ile His Tyr Ser Asn Ala Ile Ser Ser Asn Thr 405 410 415		
Thr Glu Ser Asp Lys Lys Tyr Gly Ala Ser Ser Glu Lys Met Ala Ala 420 425 430		
Ala Tyr Ala Ala Phe Ala Asn Gly Gly Thr Tyr Tyr Lys Pro Met Tyr 435 440 445		
Ile His Lys Val Val Phe Ser Asp Gly Ser Glu Lys Glu Phe Ser Asn 450 455 460		
Val Gly Thr Arg Ala Met Lys Glu Thr Thr Ala Tyr Met Met Thr Asp 465 470 475 480		
Met Met Lys Thr Val Leu Thr Tyr Gly Thr Gly Arg Asn Ala Tyr Leu 485 490 495		
Ala Trp Leu Pro Gln Ala Gly Lys Thr Gly Thr Ser Asn Tyr Thr Asp 500 505 510		
Glu Glu Ile Glu Asn His Ile Lys Thr Ser Gln Phe Val Ala Pro Asp 515 520 525		
Glu Leu Phe Ala Gly Tyr Thr Arg Lys Tyr Ser Met Ala Val Trp Thr 530 535 540		
Gly Tyr Ser Asn Arg Leu Thr Pro Leu Val Gly Asn Gly Leu Thr Val		

09765272.012264

545	550	555	560
Ala Ala Lys Val Tyr Arg Ser Met Met Thr Tyr Leu Ser Glu Gly Ser			
565		570	575
Asn Pro Glu Asp Trp Asn Ile Pro Glu Gly Leu Tyr Arg Asn Gly Glu			
580	585		590
Phe Val Phe Lys Asn Gly Ala Arg Ser Thr Trp Asn Ser Pro Ala Pro			
595	600	605	
Gln Gln Pro Pro Ser Thr Glu Ser Ser Ser Ser Ser Ser Asp Ser Ser			
610	615	620	
Thr Ser Gln Ser Ser Ser Thr Thr Pro Ser Thr Asn Asn Ser Thr Thr			
625	630	635	640
Thr Asn Pro Asn Asn Asn Thr Gln Gln Ser Asn Thr Thr Pro Asp Gln			
645	650	655	
Gln Asn Gln Asn Pro Gln Pro Ala Gln Pro			
660	665		

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1714 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

AAATTACAAT ACGGACTATG AATTGACCTC TGGAGAAAAA TTACCTCTTC CTAAAGAGAT	60
TTCAGGTTAC ACTTATATTG GATATATCAA AGAGGGAAAA ACGACTTCTG AGTCTGAAGT	120
AAGTAATCAA AAGAGTTCAG TTGCCACTCC TACAAAACAA CAAAAGGTGG ATTATAATGT	180
TACACCGAAT TTTGTAGACC ATCCATCAAC AGTACAAGCT ATTCAGGAAC AAACACCTGT	240
TTCTTCAACT AAGCCGACAG AAGTTCAAGT AGTTGAAAAA CCTTCTCTA CTGAATTAAT	300
CAATCCAAGA AAAGAAGAGA AACAATCTTC AGATTCTCAA GAACAATTAG CCGAACATAA	360
GAATCTAGAA ACGAAGAAAG AGGAGAAGAT TTCTCCAAAA GAAAAGACTG GGGTAAATAC	420
ATTAAATCCA CAGGATGAAG TTTTATCAGG TCAATTGAAC AAACCTGAAC TCTTATATCG	480
TGAGGAAACT ATGGAGACAA AAATAGATTT TCAAGAAGAA ATTCAAGAAA ATCCTGATTT	540
AGCTGAAGGA ACTGTAAGAG TAAAACAAGA AGGTAAATTA GGTAAGAAAG TTGAAATCGT	600
CAGAATATTC TCTGTAAACA AGGAAGAAGT TTCGCGAGAA ATTGTTTCAA CTTCAACGAC	660
TGCGCCTAGT CCAAGAATAG TCGAAAAAGG TACTAAAAAA ACTCAAGTTA TAAAGGAACA	720
ACCTGAGACT GGTGTAGAAC ATAAGGACGT ACAGTCTGGA GCTATTGTG AATCCGCAAT	780
TCAGCCTGAG TTGCCCGAAG CTGTAGTAAG TGACAAAGGC GAACCAGAAG TTCAACCTAC	840

09765272-012201

(2) INFORMATION FOR SEQ ID NO:4:

(A) LENGTH: 571 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Thr Glu Leu Ile Asn Pro Arg Lys Glu Glu Lys Gln Ser Ser Asp Ser
100 105 110

Gln	Glu	Gln	Leu	Ala	Glu	His	Lys	Asn	Leu	Glu	Thr	Lys	Lys	Glu	Glu
115						120						125			
Lys	Ile	Ser	Pro	Lys	Glu	Lys	Thr	Gly	Val	Asn	Thr	Leu	Asn	Pro	Gln
130						135				140					
Asp	Glu	Val	Leu	Ser	Gly	Gln	Leu	Asn	Lys	Pro	Glu	Leu	Leu	Tyr	Arg
145				150						155				160	
Glu	Glu	Thr	Met	Glu	Thr	Lys	Ile	Asp	Phe	Gln	Glu	Glu	Ile	Gln	Glu
				165				170						175	
Asn	Pro	Asp	Leu	Ala	Glu	Gly	Thr	Val	Arg	Val	Lys	Gln	Glu	Gly	Lys
		180						185				190			
Leu	Gly	Lys	Lys	Val	Glu	Ile	Val	Arg	Ile	Phe	Ser	Val	Asn	Lys	Glu
		195				200						205			
Glu	Val	Ser	Arg	Glu	Ile	Val	Ser	Thr	Ser	Thr	Thr	Ala	Pro	Ser	Pro
210						215				220					
Arg	Ile	Val	Glu	Lys	Gly	Thr	Lys	Lys	Thr	Gln	Val	Ile	Lys	Glu	Gln
225				230						235				240	
Pro	Glu	Thr	Gly	Val	Glu	His	Lys	Asp	Val	Gln	Ser	Gly	Ala	Ile	Val
				245				250						255	
Glu	Pro	Ala	Ile	Gln	Pro	Glu	Leu	Pro	Glu	Ala	Val	Val	Ser	Asp	Lys
		260						265				270			
Gly	Glu	Pro	Glu	Val	Gln	Pro	Thr	Leu	Pro	Glu	Ala	Val	Val	Thr	Asp
		275				280						285			
Lys	Gly	Glu	Thr	Glu	Val	Gln	Pro	Glu	Ser	Pro	Asp	Thr	Val	Val	Ser
290						295				300					
Asp	Lys	Gly	Glu	Pro	Glu	Gln	Val	Ala	Pro	Leu	Pro	Glu	Tyr	Lys	Gly
305				310						315				320	
Asn	Ile	Glu	Gln	Val	Lys	Pro	Glu	Thr	Pro	Val	Glu	Lys	Thr	Lys	Glu
		325						330						335	
Gln	Gly	Pro	Glu	Lys	Thr	Glu	Glu	Val	Pro	Val	Lys	Pro	Thr	Glu	Glu
		340						345				350			
Thr	Pro	Val	Asn	Pro	Asn	Glu	Gly	Thr	Thr	Glu	Gly	Thr	Ser	Ile	Gln
355						360						365			
Glu	Ala	Glu	Asn	Pro	Val	Gln	Pro	Ala	Glu	Glu	Ser	Thr	Thr	Asn	Ser
370						375				380					
Glu	Lys	Val	Ser	Pro	Asp	Thr	Ser	Ser	Lys	Asn	Thr	Gly	Glu	Val	Ser
385				390						395				400	
Ser	Asn	Pro	Ser	Asp	Ser	Thr	Thr	Ser	Val	Gly	Glu	Ser	Asn	Lys	Pro
		405						410						415	
Glu	His	Asn	Asp	Ser	Lys	Asn	Glu	Asn	Ser	Glu	Lys	Thr	Val	Glu	Glu
		420						425				430			
Val	Pro	Val	Asn	Pro	Asn	Glu	Gly	Thr	Val	Glu	Gly	Thr	Ser	Asn	Gln
		435				440						445			

Glu Thr Glu Lys Pro Val Gln Pro Ala Glu Glu Thr Gln Thr Asn Ser
 450 455 460
 Gly Lys Ile Ala Asn Glu Asn Thr Gly Glu Val Ser Asn Lys Pro Ser
 465 470 475 480
 Asp Ser Lys Pro Pro Val Glu Glu Ser Asn Gln Pro Glu Lys Asn Gly
 485 490 495
 Thr Ala Thr Lys Pro Glu Asn Ser Gly Asn Thr Thr Ser Glu Asn Gly
 500 505 510
 Gln Thr Glu Pro Glu Pro Ser Asn Gly Asn Ser Thr Glu Asp Val Ser
 515 520 525
 Thr Glu Ser Asn Thr Ser Asn Ser Asn Gly Asn Glu Glu Ile Lys Gln
 530 535 540
 Glu Asn Glu Leu Asp Pro Asp Lys Lys Val Glu Glu Pro Glu Lys Thr
 545 550 555 560
 Leu Glu Leu Arg Asn Val Ser Asp Leu Glu Leu
 565 570

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 748 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

TGAGAATCAA GCTACACCCA AAGAGACTAG CGCTCAAAAG ACAATCGTCC TTGCTACAGC	60
TGGCGACGTG CCACCATTG ACTACGAAGA CAAGGGCAAT CTGACAGGCT TTGATATCGA	120
AGTTTTAAAG GCAGTAGATG AAAAAGCTCAG CGACTACGAG ATTCAATTCC AAAGAACCGC	180
CTGGGAGAGC ATCTTCCCAG GACTTGATTG TGGTCACTAT CAGGCTGCGG CCAATAACTT	240
GAGTTACACA AAAGAGCGTG CTGAAAAATA CCTTTACTCG CTTCCAATTT CCAACAATCC	300
CCTCGTCCTT GTCAGCAACA AGAAAAATCC TTTGACTTCT CTTGACCAGA TCGCTGGTAA	360
AACAACACAA GAGGATACCG GAACTTCTAA CGCTCAATTC ATCAATAACT GGAATCAGAA	420
ACACACTGAT AATCCCGCTA CAATTAATTT TTCTGGTGAG GATATTGGTA AACGAATCCT	480
AGACCTTGCT AACGGAGAGT TTGATTTCT AGTTTTTGAC AAGGTATCCG TTCAAAAGAT	540
TATCAAGGAC CGTGGTTTAG ACCTCTCAGT CGTTGATTTA CCTTCTGCAG ATAGCCCCAG	600
CAATTATATC ATTTTCTCAA GCGACCAAAA AGAGTTTAAA GAGCAATTTG ATAAAGCGCT	660
CAAAGAACTC TATCAAGACG GAACCTTGA AAAAGCTCAGC AATACCTATC TAGGTGGTTC	720
TTACCTCCCA GATCAATCTC AGTTACAA	748

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 249 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Glu Asn Gln Ala Thr Pro Lys Glu Thr Ser Ala Gln Lys Thr Ile Val
 1 5 10 15
 Leu Ala Thr Ala Gly Asp Val Pro Pro Phe Asp Tyr Glu Asp Lys Gly
 20 25 30
 Asn Leu Thr Gly Phe Asp Ile Glu Val Leu Lys Ala Val Asp Glu Lys
 35 40 45
 Leu Ser Asp Tyr Glu Ile Gln Phe Gln Arg Thr Ala Trp Glu Ser Ile
 50 55 60
 Phe Pro Gly Leu Asp Ser Gly His Tyr Gln Ala Ala Ala Asn Asn Leu
 65 70 75 80
 Ser Tyr Thr Lys Glu Arg Ala Glu Lys Tyr Leu Tyr Ser Leu Pro Ile
 85 90 95
 Ser Asn Asn Pro Leu Val Leu Val Ser Asn Lys Lys Asn Pro Leu Thr
 100 105 110
 Ser Leu Asp Gln Ile Ala Gly Lys Thr Thr Gln Glu Asp Thr Gly Thr
 115 120 125
 Ser Asn Ala Gln Phe Ile Asn Asn Trp Asn Gln Lys His Thr Asp Asn
 130 135 140
 Pro Ala Thr Ile Asn Phe Ser Gly Glu Asp Ile Gly Lys Arg Ile Leu
 145 150 155 160
 Asp Leu Ala Asn Gly Glu Phe Asp Phe Leu Val Phe Asp Lys Val Ser
 165 170 175
 Val Gln Lys Ile Ile Lys Asp Arg Gly Leu Asp Leu Ser Val Val Asp
 180 185 190
 Leu Pro Ser Ala Asp Ser Pro Ser Asn Tyr Ile Ile Phe Ser Ser Asp
 195 200 205
 Gln Lys Glu Phe Lys Glu Gln Phe Asp Lys Ala Leu Lys Glu Leu Tyr
 210 215 220
 Gln Asp Gly Thr Leu Glu Lys Leu Ser Asn Thr Tyr Leu Gly Gly Ser
 225 230 235 240
 Tyr Leu Pro Asp Gln Ser Gln Leu Gln
 245

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 985 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

TGGTAACCGC TCTTCTCGTA ACGCAGCTTC ATCTTCTGAT GTGAAGACAA AAGCAGCAAT 60
 CGTCACTGAT ACTGGTGGTG TTGATGACAA ATCATTCAAC CAATCAGCTT GGAAGGTTT 120
 GCAGGCTTGG GGTAAAGAAC ACAATCTTTC AAAAGATAAC GGTTCACCTT ACTTCCAATC 180
 AACAAGTGAA GCTGACTACG CTAACAACCTT GCAACAAGCG GCTGGAAGTT ACAACCTAAT 240
 CTTCGGTGTT GGTTTTGCCC TTAATAATGC AGTTAAAGAT GCAGCAAAAG AACACACTGA 300
 CTTGAACTAT GTCTTGATTG ATGATGTGAT TAAAGACCAA AAGAATGTTG CGAGCGTAAC 360
 TTTCGCTGAT AATGAGTCAG GTTACCTTGC AGGTGTGGCT GCAGCAAAAA CAACTAAGAC 420
 AAAACAAGTT GGTTTTGTAG GTGGTATCGA ATCTGAAGTT ATCTCTCGTT TTGAAGCAGG 480
 ATTCAAGGCT GGTGTTGCGT CAGTAGACCC ATCTATCAAA GTCCAAGTTG ACTACGCTGG 540
 TTCATTTGGT GATGCGGCTA AAGGTAAAAC AATTGCAGCC GCACAATACG CAGCCGGTGC 600
 AGATATTGTT TACCAAGTAG CTGGTGGTAC AGGTGCAGGT GTCTTTGCAG AGGCAAAATC 660
 TCTCAACGAA AGCCGTCTCG AAAATGAAAA AGTTTGGGTT ATCGGTGTTG ATCGTGACCA 720
 AGAAGCAGAA GGTAAATACA CTTCTAAAGA TGGCAAAGAA TCAAACCTTG TTCTTGTATC 780
 TACTTTGAAA CAAGTTGGTA CAACTGTAAA AGATATTTCT AACAAGGCAG AAAGAGGAGA 840
 ATTCCTTGGC GGTCAAGTGA TCGTTTACTC ATTGAAGGAT AAAGGGGTTG ACTTGGCAGT 900
 AACAAACCTT TCAGAAGAAG GTAAAAAAGC TGTCGAAGAT GCAAAAGCTA AAATCCTTGA 960
 TGGAAGCGTA AAAGTTCCTG AAAAA 985

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 328 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Gly Asn Arg Ser Ser Arg Asn Ala Ala Ser Ser Ser Asp Val Lys Thr
 1 5 10 15
 Lys Ala Ala Ile Val Thr Asp Thr Gly Gly Val Asp Asp Lys Ser Phe
 20 25 30
 Asn Gln Ser Ala Trp Glu Gly Leu Gln Ala Trp Gly Lys Glu His Asn

09765272 012201

35					40					45					
Leu	Ser	Lys	Asp	Asn	Gly	Phe	Thr	Tyr	Phe	Gln	Ser	Thr	Ser	Glu	Ala
50						55					60				
Asp	Tyr	Ala	Asn	Asn	Leu	Gln	Gln	Ala	Ala	Gly	Ser	Tyr	Asn	Leu	Ile
65					70					75					80
Phe	Gly	Val	Gly	Phe	Ala	Leu	Asn	Asn	Ala	Val	Lys	Asp	Ala	Ala	Lys
				85					90					95	
Glu	His	Thr	Asp	Leu	Asn	Tyr	Val	Leu	Ile	Asp	Asp	Val	Ile	Lys	Asp
			100					105					110		
Gln	Lys	Asn	Val	Ala	Ser	Val	Thr	Phe	Ala	Asp	Asn	Glu	Ser	Gly	Tyr
		115					120					125			
Leu	Ala	Gly	Val	Ala	Ala	Ala	Lys	Thr	Thr	Lys	Thr	Lys	Gln	Val	Gly
		130					135					140			
Phe	Val	Gly	Gly	Ile	Glu	Ser	Glu	Val	Ile	Ser	Arg	Phe	Glu	Ala	Gly
145					150					155					160
Phe	Lys	Ala	Gly	Val	Ala	Ser	Val	Asp	Pro	Ser	Ile	Lys	Val	Gln	Val
				165					170					175	
Asp	Tyr	Ala	Gly	Ser	Phe	Gly	Asp	Ala	Ala	Lys	Gly	Lys	Thr	Ile	Ala
			180					185						190	
Ala	Ala	Gln	Tyr	Ala	Ala	Gly	Ala	Asp	Ile	Val	Tyr	Gln	Val	Ala	Gly
		195					200					205			
Gly	Thr	Gly	Ala	Gly	Val	Phe	Ala	Glu	Ala	Lys	Ser	Leu	Asn	Glu	Ser
	210					215					220				
Arg	Pro	Glu	Asn	Glu	Lys	Val	Trp	Val	Ile	Gly	Val	Asp	Arg	Asp	Gln
225					230					235					240
Glu	Ala	Glu	Gly	Lys	Tyr	Thr	Ser	Lys	Asp	Gly	Lys	Glu	Ser	Asn	Phe
				245					250					255	
Val	Leu	Val	Ser	Thr	Leu	Lys	Gln	Val	Gly	Thr	Thr	Val	Lys	Asp	Ile
			260					265					270		
Ser	Asn	Lys	Ala	Glu	Arg	Gly	Glu	Phe	Pro	Gly	Gly	Gln	Val	Ile	Val
		275					280					285			
Tyr	Ser	Leu	Lys	Asp	Lys	Gly	Val	Asp	Leu	Ala	Val	Thr	Asn	Leu	Ser
		290				295					300				
Glu	Glu	Gly	Lys	Lys	Ala	Val	Glu	Asp	Ala	Lys	Ala	Lys	Ile	Leu	Asp
305					310					315					320
Gly	Ser	Val	Lys	Val	Pro	Glu	Lys								
				325											

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1404 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

09765272.012204

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

TGTGGAAATT TGACAGGTAA CAGCAAAAAA GCTGCTGATT CAGGTGACAA ACCTGTTATC 60
 AAAATGTACC AAATCGGTGA CAAACCAGAC AACTTGGATG AATTGTTAGC AAATGCCAAC 120
 AAAATCATTG AAGAAAAAGT TGGTGCCAAA TTGGATATCC AATACCTTGG CTGGGGTGAC 180
 TATGGTAAGA AAATGTCAGT TATCACATCA TCTGGTGAAA ACTATGATAT TGCCTTTGCA 240
 GATAACTATA TTGTAAATGC TCAAAAAGGT GCTTACGCTG ACTTGACAGA ATTGTACAAA 300
 AAAGAAGGTA AAGACCTTTA CAAAGCACTT GACCCAGCTT ACATCAAGGG TAATACTGTA 360
 AATGGTAAGA TTTACGCTGT TCCAGTTGCA GCCAACGTTG CATCATCTCA AAACCTTGCC 420
 TTCAACGGAA CTCTCCTTGC TAAATATGGT ATCGATATTT CAGGTGTTAC TTCTTACGAA 480
 ACTCTTGAGC CAGTCTTGAA ACAAATCAAA GAAAAAGCTC CAGACGTAGT ACCATTTGCT 540
 ATTGGTAAAG TTTTCATCCC ATCTGATAAT TTTGACTACC CAGTAGCAAA CGGTCTTCCA 600
 TTCGTTATCG ACCTTGAAGG CGATACTACT AAAGTTGTAA ACCGTTACGA AGTGCCTCGT 660
 TTCAAAGAAC ACTTGAAGAC TCTTCACAAA TTCTATGAAG CTGGCTACAT TCCAAAAGAC 720
 GTCGCAACAA GCGATACTTC CTTTGACCTT CAACAAGATA CTTGGTTCGT TCGTGAAGAA 780
 ACAGTAGGAC CAGCTGACTA CGGTAACAGC TTGCTTTCAC GTGTTGCCAA CAAAGATATC 840
 CAAATCAAAC CAATTACTAA CTTCATCAAG NAAAACCAAA CAACACAAGT TGCTAACTTT 900
 GTCATCTCAA ACAACTCTAA GAACAAAGAA AAATCAATGG AAATCTTGAA CCTCTTGAAT 960
 ACGAACCAG AACTCTTGAA CGGTCTTGTT TACGGTCCAG AAGGCAAGAA CTGGGAAAAA 1020
 ATTGAAGGTA AAGAAAACCG TGTTGCGGTT CTTGATGGCT ACAAAGGAAA CACTCACATG 1080
 GGTGGATGGA AACTGGTAA CAACTGGATC CTTTACATCA ACGAAAACGT TACAGACCAA 1140
 CAAATCGAAA ATTCTAAGAA AGAATTGGCA GAAGCTAAAG AATCTCCAGC GCTTGGATTT 1200
 ATCTTCAATA CTGACAATGT GAAATCTGAA ATCTCAGCTA TTGCTAACAC AATGCAACAA 1260
 TTTGATACAG CTATCAACAC TGGTACTGTA GACCCAGATA AAGCGATTCC AGAATTGATG 1320
 GAAAAATTGA AATCTGAAGG TGCCTACGAA AAAGTATTGA ACGAAATGCA AAAACAATAC 1380
 GATGAATTCT TGAAAAACAA AAAA 1404

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 468 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein